What's New in SAP HANA Smart Data Integration and SAP HANA Smart Data Quality (Release Notes)
1 Overview

The SAP HANA smart data integration and SAP HANA smart data quality options provide tools to access source data, and provision, replicate, and transform that data in SAP HANA on-premise or in the cloud.

The smart data integration and smart data quality options let you enhance, cleanse, and transform data to make it more accurate and useful. These options let you efficiently connect to any source to provision and cleanse data for loading into SAP HANA on-premise or in the cloud, and for supported systems, write back to the original source.

Capabilities include:

- A simplified landscape, that is, one environment in which to provision and consume data.
- Access to more data formats including an open framework for new data sources.
- In-memory performance, which means increased speed and decreased latency.

<table>
<thead>
<tr>
<th>Feature area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAP HANA smart data integration</td>
<td>Real-time, high-speed data provisioning, bulk data movement, and federation. Provides built-in adapters plus an SDK so you can build your own.</td>
</tr>
<tr>
<td></td>
<td>Includes the following features and tools:</td>
</tr>
<tr>
<td></td>
<td>- Replication Editor in the SAP HANA Web-based Development Workbench, which lets you set up batch or real-time data replication scenarios in an easy-to-use web application</td>
</tr>
<tr>
<td></td>
<td>- Transformations presented as nodes in the application function modeler delivered with SAP HANA studio and SAP HANA Web-based Development Workbench, which lets you set up batch or real-time data transformation scenarios</td>
</tr>
<tr>
<td></td>
<td>- Data Provisioning Agent, a lightweight component that hosts data provisioning adapters, enabling data federation, replication, and transformation scenarios for on-premise or in-cloud deployments</td>
</tr>
<tr>
<td></td>
<td>- Data Provisioning adapters for connectivity to remote sources</td>
</tr>
<tr>
<td></td>
<td>- Adapter SDK to create custom adapters</td>
</tr>
<tr>
<td></td>
<td>- SAP HANA Cockpit integration for monitoring Data Provisioning Agents, remote subscriptions, and data loads</td>
</tr>
</tbody>
</table>

| SAP HANA smart data quality | Real-time, high-performance data cleansing, address cleansing, and geospatial data enrichment. Provides an intuitive interface to define data transformation flowgraphs in SAP HANA Web-based Development Workbench and SAP HANA studio. |
2 New Features in SAP HANA Smart Data Integration and SAP HANA Smart Data Quality SPS 12

Features available in SAP HANA smart data integration and SAP HANA smart data quality SPS 12.

The following topics provide you with information about the new SAP HANA smart data integration and SAP HANA smart data quality features available in SAP HANA Platform SPS 12 (Rev 120) and in the SAP HANA smart data integration Data Provisioning Agent, delivery unit, and Enterprise Semantic Services (SPS 03 Rev 00).

2.1 New Flowgraph Features

New flowgraph features and enhancements for SAP HANA smart data integration SPS 12.

The following flowgraph features and enhancements have been added to SAP HANA smart data integration SPS 12.

Data Mask node

Use the Data Mask node to protect personally identifiable or sensitive information by covering all or a portion of the data. For example, you might want to mask salary, Social Security numbers, medical identification numbers, driver’s license numbers, credit cards, and phone numbers to protect your customer or employee data from possible theft or exploitation.

Aggregation node enhancements

The Aggregation node now has filtering capabilities equivalent to the HAVING clause in SQL. You can also add, delete, rename and reorder the output columns. The output columns can be aggregate functions on the input columns.

Reload/Reconcile Data Source table

In the Data Source node, you can now reconcile differences between the source table and its structure in the node.
In the flowgraph editor, click the diff button to compare the source table with the structure in the Data Source node. Click *Reconcile* to update the node with the structure from the source table.

**Note**

Reconciling the Data Source node updates the flowgraph, but you must still save the flowgraph to make the changes permanent.

**New expression editor for Filter and Aggregation nodes**

A new expression editor is available in the Filter and Aggregation nodes to help create expressions and to validate the syntax of your expressions.

**Search for objects when building flowgraphs**

In the SAP HANA Web-based Development Workbench, you can now search for and select objects when building a flowgraph. This search capability is available for the Data Source, Data Sink, Procedure, Input Type, and Output Type nodes.

**Output unique records in Filter node**

When you have identical records from the previous node, use the Distinct checkbox in the Filter node to output unique records only. The duplicate records must match exactly; similar records will be output. For example, if the only difference between record 1 and record 2 is that the first name is spelled "Jane" and "Jayne" respectively, then both records are output.

**Automatic data type conversion**

When updating the properties of your flowgraph, there is an option to select the Data Type Conversion (for Loader only), where selecting the option automatically converts the input data type to match the mapped output column’s data type.

**Multi-select and delete in various nodes**

When deleting output columns in the Filter and Join nodes, you can multi-select using the CTRL/Shift keys.
Flowgraph runtime options

When setting flowgraph runtime options on the Properties window, previously you would choose Procedure or Task and enable Realtime in a separate option. Now there are four runtime options: Procedure, Batch Task, Realtime Task, and Transactional Task. Each of these options automatically set the realtime and batch options, as well as the task or procedure output after running the flowgraph.

Template file node

This new target node is similar to Data Sink. However, you can set up this node to use a file adapter remote source, remote object and data provisioning parameters.

History target table

You can use a Data Sink node to write to a history table by configuring parameters in the History Table Settings table.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema Name</td>
<td>The schema name of the target history table.</td>
</tr>
<tr>
<td>Name</td>
<td>The name of the target history table.</td>
</tr>
<tr>
<td>Kind</td>
<td>The type of target history table. You can choose from a Database Table or a Template Table.</td>
</tr>
<tr>
<td>Data Layout</td>
<td>The target history table layout. You can choose from COLUMN or ROW.</td>
</tr>
</tbody>
</table>

2.2 New Replication Features

New replication features and enhancements for SAP HANA smart data integration SPS 12.

The following replication features and enhancements have been added to SAP HANA smart data integration SPS 12.

Table-based replication behaviors

In prior releases, replication behavior was controlled at the replication task level, and all tables within a single replication task were handled in the same way. Now, replication is managed at the table level, and there are new replication behaviors available.
Set the replication behavior for a table by selecting the table in the Replication Task Editor and choosing the desired behavior from *Replication Behaviors*. You can set the behavior for multiple tables by selecting the tables and choosing *Set Replication Behavior*.

In addition to the *Initial load only* replication behavior, the following new behaviors are available:

**Table 1: New Replication Behaviors**

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial + realtime</td>
<td>Performs the initial load and enables realtime replication</td>
</tr>
<tr>
<td>Realtime</td>
<td>Enables realtime replication without an initial data load</td>
</tr>
<tr>
<td>No data transfer</td>
<td>Replicates only the table structure without transferring any data</td>
</tr>
<tr>
<td>Initial + realtime with structure</td>
<td>Performs the initial load, enables realtime replication, and tracks table-level changes</td>
</tr>
<tr>
<td>Realtime only with structure</td>
<td>Enables realtime replication and tracks table-level changes</td>
</tr>
</tbody>
</table>

**Truncate option for target tables in flowgraphs**

In the SAP HANA Web-based Development Workbench for the Data Sink node, previously the *Truncate* option (to clear the table before inserting data) was only available when the flowgraph is set to output a Stored Procedure. You can now also select this option for a Task Plan flowgraph.

**Remove package prefix for virtual table and target table**

When a replication task is created:

- Virtual tables are created with names in the format `<package_name>::<virtual_table_prefix><table_name>`
- Target tables are created with names in the format `<package_name>::<table_name>`

Now, you have the option to create the virtual table and target table names without the `<package_name>::` prepended to the names, using the “Use Package Prefix” option.

**Replace special characters with underscore**

In prior releases, remote object names (which might include special characters) were added to the target table name and the virtual table name. Now, all of the special characters will be replaced by an underscore in the target table names and the virtual table names.
2.3 New Monitoring and Operations Features

New monitoring and operations features and enhancements for SAP HANA smart data integration SPS 12.

The following monitoring and operations features and enhancements have been added to SAP HANA smart data integration SPS 12.

Basic usability enhancements

In this release, usability enhancements have been made to the Data Provisioning Task Monitor, Data Provisioning Agent Monitor, Data Provisioning Remote Subscription Monitor, as well as the new Data Provisioning Design Time Object Monitor.

- Data Provisioning monitor tiles in the SAP HANA cockpit now display dynamic information from each monitor.
- The default number of columns displayed in each monitor has been reduced.
  You can customize the columns that are displayed in each monitor.
- Each monitor now has a status console to display error messages, warnings, and other system information.
  You can clear the status console by right-clicking and choosing Clear Console.
- You can customize the units used for display in each monitor.
- You can easily clear any column filters in each monitor.

User settings profiles

User settings profiles are now supported in the Data Provisioning Task Monitor, Data Provisioning Agent Monitor, Data Provisioning Remote Subscription Monitor, as well as the new Data Provisioning Design Time Object Monitor.

User settings profiles allow you to quickly switch between different monitor layouts. Settings profiles contain information about visible columns, column order, column width, column filters, table visibility, and slider positions.

Email status notifications

You can now create email status notifications for various task, remote subscription, and design time object statuses:

Table 2: Supported Statuses

<table>
<thead>
<tr>
<th>Object Type</th>
<th>Supported Statuses</th>
</tr>
</thead>
</table>
| Task execution    | COMPLETED
|                   | FAILED
|                   | CANCELLED          |
Design Time Object Monitor

The Design Time Object Monitor is a new monitor that provides you with information about your design time objects, including flowgraphs and replication tasks. For example, you can see the duration of a task execution for a flowgraph and how many records have been processed.

New operations actions have been added to the Data Provisioning Design Time Object Monitor:

- Execute and schedule design time objects, including objects that require table-type parameters and variables
- View, manage, and delete design time object schedules

Agent Monitor actions

New operations actions have been added to the Data Provisioning Agent Monitor:

- Create, alter, and drop agents
- Add and remove agent locations for adapters

Remote Subscription Monitor actions and enhancements

New operations actions have been added to the Data Provisioning Remote Subscription Monitor:

- Suspend and resume capture and distribution for remote sources
- Drop and reset remote subscriptions
- Queue and distribute changes for remote subscriptions

Columns that display information about the associated design time object for a remote subscription have been added to the Remote Subscription Monitor table. You can click the name of the design time object to open it in the Design Time Object Monitor.
Task Monitor actions and enhancements

New operations actions have been added to the Data Provisioning Task Monitor:

- Start and schedule tasks, including tasks that require table-type parameters and variables
- View, manage, and delete task schedules

Columns that display information about the associated design time object for a task have been added to the Task Overview table. You can click the name of the design time object to open it in the Design Time Object Monitor.

To reduce data load, you can now limit the number of rows that are displayed in the Task Execution Monitor and Task Operation Execution Monitor tables.

2.4 New Adapter Features

New adapters and enhancements for SAP HANA smart data integration SPS 12.

The following adapters and enhancements have been added to SAP HANA smart data integration SPS 12.

See the Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality for more information.

Outlook adapter

We have added new adapter to access Microsoft Outlook source data.

Google+ adapter

We have added new adapter to access Google+ source data. You will need a Google account and you will need to obtain a Google verification code.

Camel Informix adapter

This new adapter, based on the Camel adapter, provides you with access to an Informix source, and it also allows you to write back to a virtual table.

HDFS target support in the File adapter

The File adapter now allows you to write to an HDFS target file.
Teradata adapter: New prefix and suffix options to manage object name character limitations

We have added three remote source configuration parameters (Shadow Table Prefix, Stored Procedure Suffix, and Trigger Suffix) to help you manage limitations you may have on the names of objects.

Kerberos support in the Hive adapter

We have added Kerberos as a security option in the Hive adapters.

OData adapter target support

The OData adapter now supports writing to a virtual table.

SAP HANA adapter: Support for multiple schemas and additional data type support

Prior to SPS12, when creating a remote source using the SAP HANA Adapter, you were required to specify the schema to use; thereby requiring a new remote source for each source schema. We have now relaxed this requirement.

We have also added additional data type support for the SAP HANA Adapter:

- BOOLEAN
- CHAR
- NCHAR
- BINARY

MS SQL Server Log Reader adapter: Support for new types of schema changes

We have added support for the following types of schema changes in the MS SQL Server Log Reader Adapter:

- RENAME TABLE
- RENAME COLUMN
- ALTER COLUMN
Support for MS SQL Server Log Reader adapter on Linux

You can now install and configure your systems to run a SQL Server Log Reader adapter on Linux. See the topic “Run a Microsoft SQL Server Log Reader Adapter on Linux” in the Administration Guide, for more information.

2.5 New SAP HANA Smart Data Quality Features

New features for SAP HANA smart data quality in SPS 12.

The following features are delivered for SAP HANA smart data quality in SAP HANA SPS 12.

For more information about these features, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.

New survival strategies support

In addition to the previously released Most Recent, Longest, and Priority of Sources options, you can also choose Shortest and Oldest options when creating a survival rule in the Match node.

Support for multiple survival strategies definition within a single Match node

You can create up to ten survival rule strategies in a Match node. The survival rule determines which record in a group is marked as the master record. If you create multiple survival rules, you can then prioritize the rule order to determine the master record. For example, you might have the most recent date as the highest priority followed by the longest first name.

2.6 New Enterprise Semantic Services Features

New features for the Enterprise Semantic Services component in SPS 12.

Search for remote objects in SAP HANA, BW, and ABAP remote sources

You can now publish some selected content of SAP HANA, SAP BW, and SAP ABAP remote sources using the ESS Admin Tool. This will result in extracting metadata from the remote objects that belong to the selected content of the remote sources. These extracted metadata will then be loaded into the ESS Entity Grid. Once this happens, you will be able to search for remote objects in SAP HANA, SAP BW, or SAP ABAP remote sources. For instance, you will be able to search for BEx queries.
Significant changes to the JavaScript and REST APIs

All attributes that accepted only a single valid value (such as source in publication on-demand) have been removed and the JSON structures have been simplified. In the REST API, urls and methods have changed to better follow REST guidelines, and we have exposed publication and ctid (content-type identification) APIs.

New Publisher Groups improve publication management

There are two ways to publish artifacts to the Enterprise Semantic Services Entity Grid: The HTTP REST API publish() method used by applications such as SAP HANA ADP, and the SAP HANA Enterprise Semantic Services Administration tool. If the same artifact gets published by both mechanisms, the artifact is identified in the Entity Grid Monitor as belonging to a corresponding publisher group. Therefore, publisher groups define ownership of specific publications in the Entity Grid.

When an artifact is published with a specific publisher group, it can only be unpublished by that group. If the same artifact has been published with multiple publisher groups, it can only be unpublished when all corresponding publisher groups unpublish it. This feature helps avoid conflicts between applications and an administrator using the Administration tool.

Simplified and more versatile Publish and Scheduling Monitor

The SAP HANA Enterprise Semantic Services Administration tool Publish and Scheduling Monitor interface has been redesigned. It is simpler because publication and data profiling functionalities have merged. It also has increased flexibility because you can select or deselect children artifacts for publishing independently of their parent artifacts.
3 New Features in SAP HANA Smart Data Integration and SAP HANA Smart Data Quality SPS 11

The following features are provided with SAP HANA Platform SPS 11 Rev 110 and SAP HANA smart data integration SPS 02 Rev 00.

<table>
<thead>
<tr>
<th>Features available in SAP HANA Platform</th>
<th>Features available in SAP HANA smart data integration data provisioning agent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Virtual Functions in Remote Sources</td>
<td>SOAP Adapter</td>
</tr>
<tr>
<td>Custom Parameters for Virtual Tables in Flowgraphs</td>
<td>Twitter Adapter Redesign</td>
</tr>
<tr>
<td>Schedule, Run, and Monitor Flowgraphs</td>
<td>SAP Business Suite Adapter Extractors</td>
</tr>
<tr>
<td>Load Behavior Options for Targets in Flowgraphs</td>
<td>ASE ECC Adapter</td>
</tr>
<tr>
<td>Processing Improvements</td>
<td>ASE Adapter LOB Support</td>
</tr>
<tr>
<td>Calculation View as a Reader Supports the Use of Variables</td>
<td>Microsoft Excel Adapter</td>
</tr>
<tr>
<td>Just-in-time Data Preview</td>
<td>HANA and Teradata Adapter DDL Support</td>
</tr>
<tr>
<td>Hierarchical Node</td>
<td>Camel Adapter</td>
</tr>
<tr>
<td>Smart Data Quality: Match Survival Rule</td>
<td>Enterprise Semantic Services Administration enhancements</td>
</tr>
<tr>
<td>Smart Data Quality: Additional Geocode Features</td>
<td></td>
</tr>
<tr>
<td>Smart Data Quality:: Cleanse Script Conversion</td>
<td></td>
</tr>
</tbody>
</table>

Virtual Functions in Remote Sources

In the SAP HANA Web-based Development Workbench: Catalog, you can now create virtual functions from remote sources using Data Provisioning adapters. One use case is for the SOAP adapter where virtual functions can be used to call web services.

For how to use virtual functions with Data Provisioning adapters, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.

Custom Parameters for Virtual Tables in Flowgraphs

You can create custom parameters (user-specified override parameters) to various input sources and output sources, and in this version this feature includes virtual tables. This is performed by creating an optional parameter list to be specified for a virtual table input or output source within the flowgraph XML.

For more information, see the Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.
Schedule, Run, and Monitor Flowgraphs

You can schedule and run tasks through the task monitoring, agent monitoring, or subscription monitoring user interfaces. You must have the proper permissions to apply monitoring tasks.

Load Behavior Options for Targets in Flowgraphs

For flowgraphs, you can select options that enable different target-loading behaviors and include columns that display the time and type of change made in the source. For example, you can configure the target to display all the rows that were deleted in the source, or you could include all changes to generate a change log table.

For more information, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.

Processing Improvements

This release has enhancements that allow more efficient use of resources such as cores, memory, and threads. It improves performance with reserved or limited resources (such as memory and processing limitations) when many concurrent tasks are running at the same time. This improvement uses the SAP HANA Workload Management capability to allow you to configure an optimal processing configuration.

Calculation View as a Reader Supports the Use of Variables

Calculation views are composite views that can be used to combine other views of data. A calculation view can consume other analytical views, attributes, calculation views, and tables.

This release allows the use of variables that allow data types and variables in calculation views.

Just-in-time Data Preview

Within each node in SAP HANA Web-based Development Workbench is an option to run just-in-time data preview (or JIT data preview). After enabling this option, you can run the flowgraph through the selected node, and then preview the results without running the entire flowgraph. User this feature to debug and test flowgraphs by previewing how the data is transformed after each node. The preview data is temporary and is not written to any down-stream output targets. You can see how the data is transforming when you select JIT data preview in one or more nodes in the flowgraph.

For more information, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.
Smart Data Quality: Match Survival Rule

In SAP HANA Web-based Development Workbench, you can select a survival rule that identifies the survivor record for each group of duplicate records identified through the Match node. Those records that are output with a value of "M" in the Group-Master column are considered the survivor records. You can select one of the following survivor rules.

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keep all duplicates</td>
<td>None of the records are marked as a survivor and the Group_Master column is not generated. All duplicate records in a group are output.</td>
</tr>
<tr>
<td>Length</td>
<td>The record with the largest sum of characters in one or more columns that you specify is the survivor. For example, the record with &quot;100 Main Street Ste A&quot; would survive over the record with &quot;100 Main St&quot;.</td>
</tr>
<tr>
<td>Most recent</td>
<td>The record with the latest date is the survivor. For example, the record in a Last Updated column with a date of 01/01/2016 would survive over the record with a date of 03/10/2011.</td>
</tr>
<tr>
<td>Source</td>
<td>The record with the value you prioritize is the survivor. Choose a column, and arrange the values based on priority. The record with the value that you prioritized as the highest is marked as the survivor. For example, if you consider records from the western region to be a higher priority, you would select the Region column, and then move the value &quot;West&quot; to the top. Any duplicate records with &quot;West&quot; would survive over records with a value of &quot;South&quot;, &quot;North&quot; or &quot;East&quot;. To add a source, click the + icon. To remove a source, highlight the source, and then click the X icon.</td>
</tr>
</tbody>
</table>

For more information, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.

Smart Data Quality: Additional Geocode Features

An additional mode of geocoding is available in SAP HANA Web-based Development Workbench. In addition to address geocoding, you can also perform coordinate geocoding.

<table>
<thead>
<tr>
<th>Type of geocoding</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Address</td>
<td>Where the input data contains an address, and Geocode returns the geo-location coordinates.</td>
</tr>
<tr>
<td>Coordinate</td>
<td>Where the input data contains the geo-location coordinates, and returns address data.</td>
</tr>
</tbody>
</table>

Another feature includes nearby addresses in a separate output table. You can use this feature when you select coordinate geocoding. When the latitude and longitude coordinates are input and a radius around the coordinates is set, then nearby addresses within the radius are returned in a separate output table. The separate output table includes address data and latitude and longitude data of addresses within a certain radius, additional geographic and census data (if selected), distance, assignment information, and pass-through input columns.

For more information, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.
Smart Data Quality: Cleanse Script Conversion

Use the script conversion option if you have any input data containing Chinese, Korean, or Cyrillic script and want to convert it to Latin-based characters on output.

For more information, see the *Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality*.

SOAP Adapter

This adapter is a SOAP web services client that can talk to a web service using the HTTP/HTTPS protocol to download the data. The SOAP adapter uses virtual functions instead of virtual tables to expose server side operations as it closely relates to how the operation is invoked.

Twitter Adapter Redesign

The Twitter adapter is updated with new functionality.

SAP ABAP Adapter

This adapter retrieves data from virtual tables through RFC for ABAP tables and ODP extractors. It also provides change data capture for ODP extractors.

SAP ASE ECC Adapter

This adapter retrieves data from an SAP ERP system running on SAP ASE.

SAP ASE Adapter LOB Support

LOB Support has been added to the ASE adapter.

SAP HANA and Teradata Adapter DDL Support

DDL propagation support has been added the SAP HANA and Teradata Adapters.
Microsoft Excel Adapter

This adapter retrieves data from MS Excel.

Camel Adapter

The Camel adapter is a framework where you can create custom adapters using a few configuration files and Spring DSL (Domain Specific Language) with little or no coding effort. It is based on Apache Camel and the Spring Framework. To use the Camel adapter, you must know Apache Camel and Spring Framework.

Camel Facebook Adapter

The Camel Facebook adapter is a predelivered component that is based upon Camel adapter. Use the Camel Facebook component to connect to and retrieve data from Facebook.

In addition to Facebook, many other components are available from the Apache Software Foundation website.

Hierarchical Node

The Hierarchical Data Provisioning node accepts hierarchical data such as an XML schema and flattens it to one or more outputs. This can be used in combination with the SOAP adapter to convert the response of a web service function into a flat structure.

For how to configure the node, see the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.

Enterprise Semantic Services Administration Enhancements

The SAP HANA ESS Administration tool has been significantly enhanced as follows:

- Publish and unpublish artifacts (new)
- Schedule publication and data profiling tasks (new)
- Monitor blacklisted objects (new)
- Monitor all requests (new)
- Monitor the Entity Grid (enhanced)

For details, see the Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality.
4  

New Features in SAP HANA Smart Data Integration and SAP HANA Smart Data Quality SPS 10 DSP (Rev 102.1) and SAP HANA Smart Data Integration SPS 01 Rev 01

The following features are provided with SPS 01 Rev 01 and SAP HANA Platform SPS 10 Rev 102.1.

Data Provisioning Agent Grouping

Agent grouping provides fail-over and load balancing capabilities by combining multiple Data Provisioning Agents installed on separate host systems into agent groups.

For information about how to configure agent grouping in your system landscape, see the Administration Guide.

IBM DB2 for z/OS Adapter

The new IBM DB2 for z/OS adapter provides access to data from an IBM DB2 database deployed on an IBM z/OS mainframe. This adapter also allows you to write to a virtual table using a Data Sink node in a flowgraph.

Data preview in the SAP HANA Web-based Development Workbench

You can preview the data in the Data Source, Template Table and Data Sink nodes in the SAP HANA Web-based Development Workbench. Choose the Data Preview icon on the canvas to open the data in a separate tab where you can verify the data and column structure.

Note

Data preview in the Data Sink node shows existing data (if any) in that table or view. It does not show how the data might appear after being transformed in the flowgraph.
Auto-arrange the flowgraph in SAP HANA Web-based Development Workbench

Use the auto-arrange feature to automatically place the nodes in order on the canvas. This may help in tracking the flow of data to the output table.

Calculation view as a reader

You can create a graphical or scripted (using SQLScript) calculation view of your data, and then select that calculation view from the catalog as a data source within a flowgraph. The calculation view therefore serves as the data reader.

**Note**

A calculation view does not allow the use of variables and parameters.
New Features in SAP HANA Smart Data Integration and SAP HANA Smart Data Quality SPS 10 (Rev 102)

The following smart data quality feature has been delivered for the SAP HANA SPS 10 Rev 102 release.

Geocode node in the SAP HANA Web-based Development Workbench

The Geocode node is available in SAP HANA Web-based Development Workbench. Latitude and longitude data is produced based on the best input address-level assignment using the information available in the address reference data. You can use latitude and longitude data in a variety of ways such as listing business that are near an address, making targeted marketing campaigns, and so on. The data type for these columns is ST_POINT.
In SAP HANA smart data integration and SAP HANA smart data quality SPS 10, new features and enhancements are provided for administration, development, remote source connection, monitoring, and others.

**New Data Provisioning Adapters**

We have added three new pre-delivered data provisioning adapters: SAP ASE, Teradata, and SAP HANA. For information about these adapters and the features available with these adapters, see the *Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality*.

**Virtual tables as targets**

For some remote sources, using data provisioning adapters, you can now write back to a virtual table using the Data Sink node in a flowgraph. The following adapters currently support this functionality:

- Log Reader adapters (not including the ECC adapters)
- SAP ASE
- SAP HANA
- Teradata
- File

**Searching a Remote Source**

In the SAP HANA Web-based Development Workbench Catalog, you can search remote sources to find objects and create virtual tables. See the *Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality* for more information.

**Propagation of DDL changes**

DDL changes from remote sources are propagated to SAP HANA in real time.

For example, a schema change such a new column added to the remote source database can be automatically propagated to SAP HANA so that the corresponding virtual table, remote object, and target table are altered.
accordingly where the remote subscription uses table-to-table replication. This is done without requiring a restart of the system.

DDL propagation is supported by only certain adapters. See the adapter documentation in the Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality for more information.

**Improved performance of batch loads**

For better performance, you can now create logical partitions that will let the system perform parallel queries on a virtual table. These logical partitions can be created in a replication task or a flowgraph (in the Data Source Node).

**Profiling data**

Data profiling provides tools to perform a data quality assessment by examining existing data (including virtual tables) to collect statistics and information about that data. Data profiling is performed automatically in the Development Workbench to assist in determined proper input field mappings.

The three types of data profiling available are semantic profiling, distribution profiling, and metadata profiling. The profiling functionality can be called directly from readily available built-in stored procedures.

**SAP HANA Web-based Development Workbench**

In this release, the SAP HANA Web-based Development Workbench has been expanded from only replicating data to transforming data by creating flowgraphs. This intuitive interface assists you in setting up data quality nodes for both Match and Cleanse. All other data integration nodes are also available, such as join, filter and search. With the SAP HANA Web-based Development Workbench, you can use the same tool for replicating and transforming your data. Most of the nodes available in application function modeler within SAP HANA studio are also available in the SAP HANA Web-based Development Workbench. The Match node is only available in this tool. With the Match node, you can identify potentially duplicate (matching) records.

For both the Cleanse node and the Match node, a wizard guides you through the setup process. The wizards present logical choices, which you can customize to meet your needs:

- In Match, you can determine which components are matched on, for example, Address or Person, and how closely they must match.
- In Cleanse, you can choose the components you want to cleanse, assign content types, select additional columns to output, and select the format of the output data.

See the Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality for details.
Geocode Node

The Geocode node generates a new Latitude and Longitude output column, which shows the latitude and longitude of the input address at the best level that the address can be assigned to the reference data. The data type for this column is ST_POINT.

**Note**

Currently, the Geocode node is only available in the application function modeler of SAP HANA studio.

New and Updated SQL Statements

Refer to the SAP HANA SQL Reference and System Views Reference for complete information about these SQL statements.

Table 3: SQL Statements

<table>
<thead>
<tr>
<th>SQL Statement</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALTER REMOTE SOURCE</td>
<td>Now includes CLEAR OBJECTS, REFRESH OBJECTS, and CANCEL REFRESH OBJECTS</td>
</tr>
<tr>
<td>CANCEL TASK</td>
<td>New SQL statement that stops the execution of a task plan</td>
</tr>
<tr>
<td>CREATE REMOTE SUBSCRIPTION</td>
<td>Now includes syntax changes to support procedure parameters when the target in a remote subscription is a task</td>
</tr>
<tr>
<td>SESSION CONTEXT</td>
<td>Now returns the value of the session_variable assigned to the current user</td>
</tr>
<tr>
<td>START TASK</td>
<td>Now returns a TASK_EXECUTION_ID session variable, a unique task execution ID</td>
</tr>
</tbody>
</table>

New and Updated System Views and Monitoring Views

Refer to the SAP HANA SQL Reference and System Views Reference for complete information about these views.

Table 4: System Views and Monitoring Views

<table>
<thead>
<tr>
<th>View</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>M_AGENTS</td>
<td>Now includes agent version</td>
</tr>
<tr>
<td>M_SESSION CONTEXT</td>
<td>Now shows task execution ID</td>
</tr>
<tr>
<td>REMOTE_SUBSCRIPTION</td>
<td>Now shows TASK_PROCEDURE_PARAMETERS, a comma-separated list of task parameters</td>
</tr>
<tr>
<td>TASK_OPERATIONS</td>
<td>New view that shows operations-level task statistics generated when START TASK is run</td>
</tr>
<tr>
<td>TASK_OPERATIONS_EXECUTIONS</td>
<td>New view that shows task-level run-time statistics generated when START TASK is run</td>
</tr>
</tbody>
</table>
### View | Change
--- | ---
**TASKS** | Now shows:

- **TASK_OID**, unique identifier for a task
- **OWNER_NAME**, owner of the task
- **HAS_SDQ**, indicates if the task contains functionality
- **SQL_SECURITY**, security model for the task

---

### Side-effect Data

Side-effect data is additional information and statistics about data processing performed by flowgraphs and nodes such as Cleanse and Match, and provides further insight into what the process actually did. You can use this information to debug problems in your flowgraph, isolate performance bottlenecks, and so on.

For example, in addition to operational statistics about the flowgraph itself, side-effect data generated by a Cleanse node might include statistics about how many addresses were transformed as well as information about the ways that the addresses were modified or enhanced.

For more details about configuring the side-effect data available in the SPS 10 release, see the “Cleanse” and “Match” sections of the *Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality*.

### Data Provisioning Agent Upgrade

In this release, the Data Provisioning Agent installation program has been enhanced with a new upgrade mode for upgrading existing agents to a new major version.

For complete information about upgrading the Data Provisioning Agent, see the *Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality*.

### Enterprise Semantic Services

This version includes a new component, Enterprise Semantic Services, which provides the following capabilities:

- **Enterprise Semantic Services** lets you search for publication artifacts or run-time objects based on their metadata and contents. For more information about Enterprise Semantic Services, see the *Administration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality* and the *Configuration Guide for SAP HANA Smart Data Integration and SAP HANA Smart Data Quality*.
- **Javascript Application Programming Interfaces** provide the following services:
  - **Search and Suggestion Service**: Provides semantic search API to search EntitySets in the Entity Grid by matching indexed metadata and data values, plus a keyword Suggestion API to provide keyword suggestions based on partial keywords you type in real time.
  - **Publication Service**: Used for managing the contents of the Entity Grid through publications.
○ Content-Type Identification Service: Lets you retrieve the content types of given attributes belonging to EntitySets in the Entity Grid and perform on-demand content-type analysis of columns in tables, SQL views, or SAP HANA views that are not yet known to the Entity Grid.

● An HTTP REST API is provided for the Search and Suggestion Service. For more information about how to use these APIs, see the SAP HANA Enterprise Semantic Services JavaScript and REST API References.

Load Behavior Options for Target Tables

For replication tasks, you can now select the way target tables load. You can select from the following options on the task’s Load Behavior tab:

● Replicate: Replicates insert, update and delete operations
● Replicate with logical delete: Replicates insert and update operations and converts delete operations to updates
● Preserve all: Replicates insert operations and converts update and delete operations to insert operations, resulting in a history table that contains all changes over time.
7 Supported Data Provisioning Nodes and Adapters

SAP HANA smart data integration and SAP HANA smart data quality provide many application function modeler nodes for data provisioning and data quality, as well as many out-of-the-box adapters for accessing source data.

Supported Data Provisioning Nodes

The following nodes are supported in SAP HANA studio (application function modeler) or the SAP HANA Web-based Development Workbench’s flowgraph editor:

<table>
<thead>
<tr>
<th>Node</th>
<th>Description</th>
<th>Where supported (SAP HANA studio, Workbench)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AFL Function</td>
<td>Accesses functions of the Application Function Library.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Aggregation</td>
<td>Creates a new aggregated result set using the user-specified aggregation method on one or more columns of data. The aggregation methods supported are: count, min, max, sum, avg and group-by.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Case</td>
<td>Routes records to one or more output paths.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Cleanse</td>
<td>Parses, standardizes, and enriches person, title, phone, firm, email, and address information within a specified input source.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Date Generation</td>
<td>Creates a one column result set with a user-defined number of rows specific to date by incrementing the starting date daily, weekly, or monthly.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Filter</td>
<td>Creates a result set from a single input source to remove rows of data using the specified SQL expression.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Geocode</td>
<td>Enriches address data with associated latitude and longitude information.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>History Preserving</td>
<td>Allows for maintaining older versions of rows when a change occurs by generating new rows in a target table with the changed values, instead of updating existing rows. This operation will use the OP_CODE column’s value of an input row to determine the result set.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Join</td>
<td>Used to combine rows from two input tables based on common attributes between them.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Lookup</td>
<td>Retrieves a column value or values from a lookup table that matches a user specified expression. User can sort result list when multiple rows are returned or configure default values in the form of constants when no rows are returned.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Map Operation</td>
<td>Modifies data based on mapping expressions and current operation codes. The operation codes can be converted between data manipulation operations.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Node</td>
<td>Description</td>
<td>Where supported</td>
</tr>
<tr>
<td>-----------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------------------------------------------</td>
</tr>
<tr>
<td>Match</td>
<td>Identifies potentially duplicate (matching) records. Match analyzes data, identifies content types and match components, and recommends match policies. You can accept the recommended policies or choose different policies. You can also adjust match settings, which control special scenarios.</td>
<td>Workbench only</td>
</tr>
<tr>
<td>Pivot</td>
<td>Transforms rows into columns of a specified input source. Any attribute that will be pivoted will need to contain a set of values that are to be transformed into columns.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Procedure</td>
<td>Calls a user-defined stored procedure from within a transformation flow</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Row Generation</td>
<td>Creates a one column result set with a user-defined number of rows specific to integer values by incrementing the starting value by 1 and ending at a user defined max.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Sort</td>
<td>Creates a result set from a single input source using the specified SQL expression sort criteria.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Table Comparison</td>
<td>Compares two input sources and produces the difference between them as a result set with rows flagged as insert, update or delete. The OP_CODE column will be used to identify records that were inserted, updated, or deleted to sync the comparison table with the input table.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Union</td>
<td>Produces a result set of two input sources of the same schema that represents either distinct rows (union) or non-distinct rows (union all).</td>
<td>SAP HANA studio, Workbench</td>
</tr>
<tr>
<td>Unpivot</td>
<td>Transforms columns into rows of a specified input source.</td>
<td>SAP HANA studio, Workbench</td>
</tr>
</tbody>
</table>

**Supported Adapters**

The following adapters are supported:

<table>
<thead>
<tr>
<th>Adapter</th>
<th>Data Source Description</th>
<th>Available as of version</th>
</tr>
</thead>
<tbody>
<tr>
<td>IBM DB2 Log Reader</td>
<td>This adapter retrieves data from DB2 and can write back to a virtual table. It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>Oracle Log Reader</td>
<td>This adapter retrieves data from Oracle and can write back to a virtual table. It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>MS SQL Server Log Reader</td>
<td>This adapter retrieves data from MS SQL Server and can write back to a virtual table. It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>DB2 ECC</td>
<td>This adapter retrieves data from an SAP ERP system running on DB2. It can also receive changes that occur to tables in real time. The only difference between this adapter and the DB2LogReaderAdapter is that this adapter uses the data dictionary in the SAP ERP system when browsing metadata.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>Adapter</td>
<td>Data Source Description</td>
<td>Available as of version</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Oracle ECC</td>
<td>This adapter retrieves data from an SAP ERP system running on Oracle. It can also receive changes that occur to tables in real time. The only difference between this adapter and the OracleLogReaderAdapter is that this adapter uses the data dictionary in the SAP ERP system when browsing metadata.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>MS SQL Server ECC</td>
<td>This adapter retrieves data from an SAP ERP system running on SQL Server. It can also receive changes that occur to tables in real time. The only difference between this adapter and the MssqlLogReaderAdapter is that this adapter uses the data dictionary in the SAP ERP system when browsing metadata.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>IBM DB2 for z/OS</td>
<td>This adapter retrieves data from DB2 for z/OS and can write back to a virtual table.</td>
<td>SPS 09 (Rev 102), SDI SPS 01 Rev 01</td>
</tr>
<tr>
<td>File</td>
<td>This adapter retrieves data from formatted and unformatted text files.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>Hive</td>
<td>This adapter retrieves data from HADOOP.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>OData</td>
<td>This adapter retrieves data from an OData service.</td>
<td>SPS 09</td>
</tr>
<tr>
<td>SAP HANA</td>
<td>This adapter retrieves data from SAP HANA and can write back to a virtual table (realtime only). It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 10</td>
</tr>
<tr>
<td>SAP ASE</td>
<td>This adapter retrieves data from SAP ASE and can write back to a virtual table (realtime only). It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 10</td>
</tr>
<tr>
<td>Teradata</td>
<td>This adapter retrieves data from Teradata and can write back to a virtual table (realtime only). It can also receive and pass along changes that occur to tables in real time.</td>
<td>SPS 10</td>
</tr>
<tr>
<td>Twitter</td>
<td>This adapter retrieves data from Twitter. It can also receive new data from Twitter in real time.</td>
<td>SPS 10</td>
</tr>
<tr>
<td>MS Excel</td>
<td>This adapter retrieves data from MS Excel.</td>
<td>SPS 11</td>
</tr>
<tr>
<td>SAP ASE ECC</td>
<td>This adapter retrieves data from an SAP ERP system running on SAP ASE.</td>
<td>SPS 11</td>
</tr>
<tr>
<td>Camel Facebook</td>
<td>The Camel Facebook adapter is a pre-delivered component that is based upon Camel adapter. Use the Camel Facebook component to connect to and retrieve data from Facebook. In addition to Facebook, many other components are available from the Apache Software Foundation website.</td>
<td>SPS 11</td>
</tr>
<tr>
<td>SAP ABAP</td>
<td>This adapter retrieves data from virtual tables through RFC for ABAP tables and ODP extractors. It also provides real-time change data capture for ODP extractors.</td>
<td>SPS 11</td>
</tr>
<tr>
<td>SOAP</td>
<td>This adapter is a SOAP web services client that can talk to a web service using the HTTP protocol to download the data. The SOAP adapter uses virtual functions instead of virtual tables to expose server side operations as it closely relates to how the operation is invoked.</td>
<td>SPS 11</td>
</tr>
<tr>
<td>Camel Informix</td>
<td>The Camel Informix adapter is a pre-delivered component that is based upon Camel adapter. This adapter retrieves data from an Informix source. It can also write back to an Informix virtual table.</td>
<td>SPS 12</td>
</tr>
</tbody>
</table>
### Related Information

**Product Availability Matrix for SAP HANA Smart Data Integration**

<table>
<thead>
<tr>
<th>Adapter</th>
<th>Data Source Description</th>
<th>Available as of version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google+</td>
<td>This adapter retrieves data from Google+.</td>
<td>SPS 12</td>
</tr>
<tr>
<td>Microsoft Outlook</td>
<td>This adapter retrieves data from Microsoft Outlook.</td>
<td>SPS 12</td>
</tr>
</tbody>
</table>
Important Disclaimers and Legal Information

Coding Samples

Any software coding and/or code lines / strings ("Code") included in this documentation are only examples and are not intended to be used in a productive system environment. The Code is only intended to better explain and visualize the syntax and phrasing rules of certain coding. SAP does not warrant the correctness and completeness of the Code given herein, and SAP shall not be liable for errors or damages caused by the usage of the Code, unless damages were caused by SAP intentionally or by SAP's gross negligence.

Accessibility

The information contained in the SAP documentation represents SAP's current view of accessibility criteria as of the date of publication; it is in no way intended to be a binding guideline on how to ensure accessibility of software products. SAP in particular disclaims any liability in relation to this document. This disclaimer, however, does not apply in cases of wilful misconduct or gross negligence of SAP. Furthermore, this document does not result in any direct or indirect contractual obligations of SAP.

Gender-Neutral Language

As far as possible, SAP documentation is gender neutral. Depending on the context, the reader is addressed directly with "you", or a gender-neutral noun (such as "sales person" or "working days") is used. If when referring to members of both sexes, however, the third-person singular cannot be avoided or a gender-neutral noun does not exist, SAP reserves the right to use the masculine form of the noun and pronoun. This is to ensure that the documentation remains comprehensible.

Internet Hyperlinks

The SAP documentation may contain hyperlinks to the Internet. These hyperlinks are intended to serve as a hint about where to find related information. SAP does not warrant the availability and correctness of this related information or the ability of this information to serve a particular purpose. SAP shall not be liable for any damages caused by the use of related information unless damages have been caused by SAP's gross negligence or wilful misconduct. All links are categorized for transparency (see: http://help.sap.com/disclaimer).